Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 46 (2015) 176 – 183
Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

Efficient Knowledge Transformation System Using Pair of Classifiers for Prediction of Students Career Choice

Roshani Ade [a,*], P. R. Deshmukh[a]

[a]*Department of Computer Science and Engineering, Sipna College of Engineering and Technology, SGBAU, Amravati, 444607, India*

Abstract

The ability to predict the career of students can be beneficial in a huge number of different techniques which are connected with the education structure. Student's marks in psychometric test can form the training set for the system which helps the students to choose the right career. As the student's data in the educational systems is increasing day by day, the incremental learning properties are important for machine learning research. Against to the classical batch learning algorithm, incremental learning algorithm tries to forget unrelated information while training new instances. Effective knowledge transformation system can be build using different pair of classifiers for the purpose of prediction of student's career choice. In this paper, four pair of classifier are used.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Incremental Learning; Education Structure; Knowledge Transformation; Concept Class; Pair Of Classifiers

1. Introduction

Now a day's choosing right career is one of the most important aspects of the students learning process, and it is difficult to choose the right career option when the plenty of options are available. The important aspects like interest, talent with some psychological parameters like self-awareness, empathy, self-motivation, emotional stability, relation management, truthfulness etc. are important to consider before choosing a right career path.

* Corresponding author. Tel.: +91-9960866438.
E-mail address: rosh513@gmail.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.009

It is commonly seen that, many of the students have their poor academic record just because of choosing their career without considering their own capabilities, consequently which leads to their waste of time and the money, so it is important to choose the right career in the first place of the career path. For the purpose of the same, the psychometric test can be conducted on the students and the students are classified for choosing their best career option.

In the supervised machine learning strategies, the model is built using predefined classes which produces the model hypothesis, then the hypothesis is used to predict the unseen instances which are new to the system. In the simple batch learning system[1,2,3] using supervised learning algorithms like MLP, SVM, NB, CART, C4.5, LVQ etc. which needs to be retrained when new data introduces with the system and the system is not able to detect new class. In this study, the new class of students is nothing but the student having different features introduces after some time period or introduces in the next chunk of data. The student's classification systems in the literature are used to assess the of student's performance by using a neuro-fuzzy classification technique[4], for the classification of educational background of students by using genetic neural network[5]. The work in the literature for the classification of students has not focused on prediction of student's career choice and incremental learning technique is not used for purpose of the same. Incremental learning algorithm has ability to learn from new incoming data to the system even after the classifier has already been generated from the previously available data. This concept is used for various applications like autonomous navigation system[6,7], organization of emails[8], diagnosing faults in the nuclear power system[9] and many other using the techniques like SVM[10,11,12,13,14,15], fuzzy neural network16, incremental genetic algorithms and others[17,18,19] in the literature. For example, student's carrier preferences changes as new courses or facilities become available. So, the learning algorithm needs to obtain the set of concept explanation from training data which was scattered over time. Algorithms for copying concept drift, definition of a class is changing as the time passes, must congregate speedily and precisely to new goal concept and it should be efficient in terms of space and time complexity. The model used in this paper, for incremental learning of students uses the pair of classifiers for different experiments. The first pair of the classifier is used for the base classifier in the model and the second classifier in the model is used for weight distribution function. The base classifier CART and SVM are used for the classification of the data and the MLP and the SVM is used for the weight distribution function. In this study, conceptual view of the system is designed and experimental results on the student's data as well as some real world data sets are used to prove the potency of the proposed method.

This paper is ordered as follows, section 2 apprises about prediction of student's career choice, and introduces incremental learning techniques in the literature and the applications of incremental learning concept. Section 3 explain the proposed algorithm, section 4 gives the experiments and results of the proposed algorithm student's data. Finally section 5 gives the conclusion of the work and its future scope.

2. Background and Related Work

In this section the overview of incremental learning and the concept behind pair of classifiers is given.

2.1. Incremental Learning

A practical approach for learning from new data is nothing but, discarding old classifier and again training the new classifier, with all the data. This type of approach is having the problem of catastrophic forgetting. It is not a desirable approach as retraining is involved which is financially costly as well as it is not possible to train the classifier if the original dataset is lost. Efficient incremental learning algorithm can be designed if the algorithm achieves all the properties of incremental learning mentioned below.

- While training, it should require small stable time per sample
- There should be only one sample at a time in memory, so the fixed amount of memory will be used
- It will build the model by just scanning the database only once.
- It should preserve previously obtained knowledge.

The incremental learning algorithm produces sequences of hypotheses for the sequence of training samples, where recent hypothesis is nothing but the description of all data which have been accumulated thus far and it always hinge on preceding hypothesis and the recent training data[20,21,22,23]. Incremental learning algorithm should learn the novel information and it should preserve formerly acquired knowledge without accessing the formerly seen data so far.

In the real time application, it is very much common for the data to appear in batches, like the students data appears yearly, we can say a batch of a year. To handle this incremental batch of data, there is a need to transformation of knowledge to old batch to the current batch or from current batch to new batch. If new data contains new class of a student, the system will be able to detect that new class without degradation in performance.

The use of ensemble technique in incremental learning is not uncommon, which is needed to ensemble the output of different base classifiers. The algorithm20 focuses on the four properties of incremental learning using the concept of ADABOOST and the majority voting rules of the ensemble is used to combine the output of classifiers. There are mainly ten rules[24,25,26,27,28,29,30] with the addition of SSC[31] which can be used to combine the output of classifiers to get the final hypothesis. The algorithms used in the literature for incremental learning uses the ensembling two times or we can call it as double ensembling strategy. Learn++ uses the double ensembling strategy, it creates the ensemble of weak classifiers, and each trained on the subset of original datasets. At each iteration the samples which are difficult to classify are given more weightage for the classification. Because of the double ensemble strategy the time needed for the computation is more as the number of hypothesis are drawn from each data chunk. Instead of number of hypotheses, our proposed approach draw only one hypothesis from each data chunk, so there will be only single ensemble strategy is used as described in the next section.

2.2. Pair of Classifiers

In the pair of classifiers concept, as described earlier, the combinations of base classifiers and classifiers for the purpose of weight distribution are used. The four pairs of classifiers namely CART-MLP, CART-SVM, SVM-MLP, SVM-SVM are formed for the experiments.

3. Proposed Algorithm

Fig. 1. gives the description of the algorithm. The input to the algorithm is data chunk coming at time t. Here, we are considering a data coming at time t and the data coming at time t+1, called as Dt and Dt+1 respectively. Each data chunk is having its weight distribution along with it called as DFt and DFt+1 respectively. The overall classification has been done with the pair of classifiers. The first classifier in the pair is CART or SVM, which is used as a base classifier for classification. The ht is nothing but hypothesis generated by the base classifier. The second classifier in the pair MLP or SVM is used for calculation of a weight distribution. The input to the MLP is Dt and DFt, after the stabilization of MLP, this weight will be used as a weight distribution for Dt+1.The error will be calculated by applying ht on Dt+1 with DF't weight distribution function. Then β gives normalized error. In the next step of the algorithm, the weights will be updated, so that the samples which are misclassified will get more importance in the next iteration. In this way the same process will be repeated for the next coming data chunks. Then the final hypothesis will be combined by using weighted majority voting rule, given in step 9 of Fig. 1.

3.1. Dataset Used

The dataset is created by conducting psychometric test on 1333 students of age group 16 to 20. The dataset contains 1333 samples, 14 attributes, out of 14, 10 attributes are as described in Table 1 and 4 attributes are nothing but the marks rank in the four subjects like subject1 to subject4 and 7 classes, which denotes the categories of the interest of the student[32,33]. The attributes are shown in Table 1. All the attributes are numeric attributes, except the

Step 1: Each data set is initially sliced into D data chunks (any value of D can be chosen, which specify the chunk size) $\{D_1, D_2, \ldots, Dt\}$, No. of datasets= T_k

Step 2: Initialize the weights DF_t for the first dataset, $DF_t = \{w1, w_2, \ldots, w_t\}$, which is uniform for the first chunk, as nothing has been learned yet.

Step 3: Call base classifier (CART or SVM) with weight distribution DF_t, which generates the hypothesis h_t

Step 4: Weights are updated (called as DF'_t) using second classifier in the pair (MLP or SVM) which takes the input as D_t,DF_t, once the model is stabilized, updated weight DF_{t+1} is used as a weight distribution function for D_{t+1}.

Step 5: The h_t will be applied on D_{t+1} with DF'_t, and calculates error using following formula

$$\varepsilon_t = \sum_{i:ht(xi)=yi} DF't(i)$$

Step 6: Normalization of the error has been done using

$$\beta_t = \varepsilon_t / (1 - \varepsilon_t)$$

Step 7: Normalize weight distribution using following equation, misclassified examples weight will not get change

$$DF'_{t+1} = \frac{DF't}{Zt} \text{ X } \begin{cases} \beta t, \text{ if } ht(xi) = yj \\ 1, \text{otherwise} \end{cases}$$

Step 8: Repeat the step3 to Step7 for each data chunk

Step 9: Get the final hypothesis using weighted majority voting rule.

$$xt \rightarrow yj \text{ satisfy } \max_{yj} \sum_{i=1}^{L} wi.\, \Delta i\left(\frac{yj}{xt}\right)$$

Fig. 1. Proposed Algorithm

Table 1. Attributes of Students Dataset.

A	B	C	D	E	F	G	H	I	J
Self Awareness	Empathy	Self Motivation	Emotional Stability	Managing Relations	Integrity	Self Development	Value Orientation	Commitment	Altruistic Behavior

class attribute. The class value depends on total score of a student. The total score is the addition of a score of all the 14 attributes and there are 7 classes of the student. The same algorithm is also applied to the turkey dataset available on UCI repository is used[34]. Total samples in Turkey dataset are 5020, having total 32 attributes and 3 classes. All the attributes in Turkey Students dataset are numeric attributes.

4. Experiment Results

For the intention of this study, the data set of students which has been created having attributes shown in Table 1 and the turkey dataset have been used. To calculate the classification accuracy, the entire training set was divided into ten mutually exclusive and equal sized sets, i.e. 10 cross validation is used

4.1 Experiments with pair of Classifiers

In the practical implementation each data set is divided into10 data chunks. Out of 10 chunks in each iteration randomly one chunk is selected for the testing and the remaining 9 chunks are fed to the proposed algorithm shown in Fig. 1. In the first experiment of this study, the CART-MLP pair of classifier is used, CART is used as a base classifier and the MLP is used for weight distribution. The MLP parameters like epochs are set to 50, input layer is having 14 neurons for 14 attributes, and one output neuron and the 5 hidden layer neurons. In the same experiment SVM-MLP pair is used, in this SVM is used as a base classifier an MLP is used for weight distribution with same settings. The prediction accuracy of the same is shown in Fig. 2. In the second experiment, CART-SVM an SVM-SVM pairs are used, and the result s of the same are shown in Fig. 3.

The class wise accuracy of all the methods are shown in Table 2 and the comparison of the same is given with Learn++.NC technique.

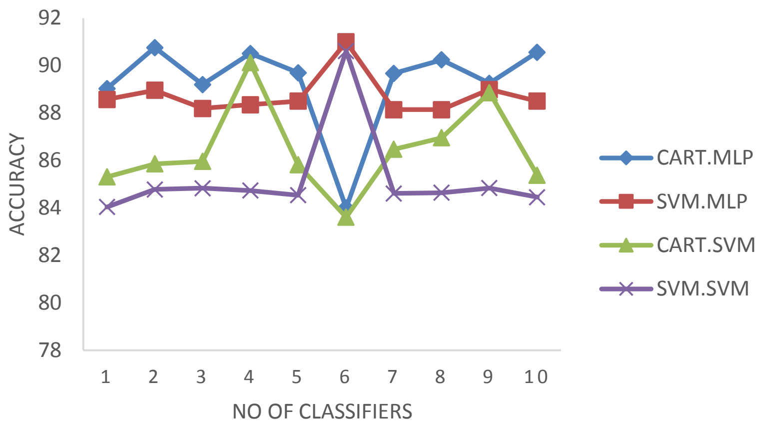

Fig. 2. Prediction accuracy of student's dataset with CART or SVM as a base classifier and MLP or SVM for weight distribution

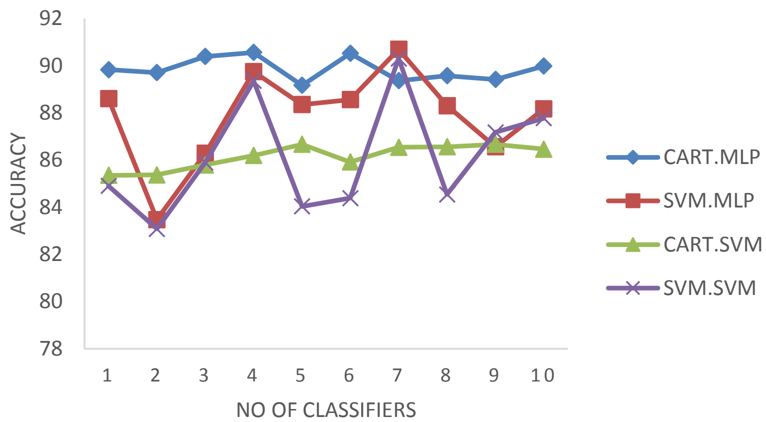

Fig. 3. Prediction accuracy of Turkey student's dataset with CART or SVM as a base classifier and MLP or SVM for weight distribution

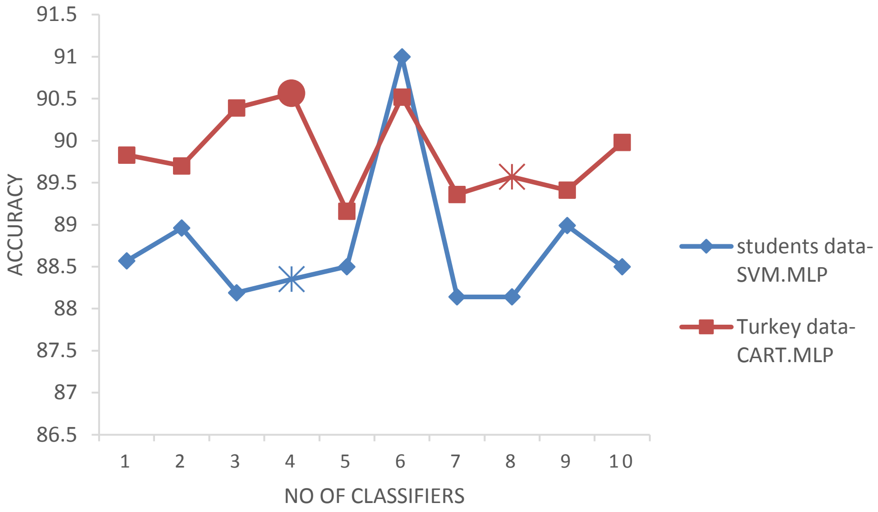

Fig. 4. Single and multiple concept class detection

Table 2. Class wise accuracy

Dataset	Methods	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Overall Accuracy
Students Dataset	Learn++.NC	83.48	82.12	83.88	81.15	81.11	80.09	82.11	82.06
	CART.MLP	90.75	89.71	90.45	89.62	90.1	90.97	89.26	90.69
	SVM.MLP	88.49	88.59	88.22	88.35	88.84	88.25	88	90.2
	CART.SVM	89.2	89.16	89.65	89.18	89.03	89.53	89.61	90.52
	SVM.SVM	85.5	85.98	85.1	85.14	85.37	85.59	84.67	82.06
Turkey Students Dataset	Learn++.NC	83.26	82.12	83.45					83.78
	CART.MLP	90.84	89.71	89.11					89.07
	SVM.MLP	88.47	99.74	99.65					88.16
	CART.SVM	90.98	89.83	90.34					90.85
	SVM.SVM	85.92	99.74	99.65					88.77

4.2 Concept Class Detection

Detecting a new class which comes with new data is one of the property of incremental learning. It might happen that, the new chunk of data is having the samples having new class. Then the system should able to detect the new class efficiently. For the simulation of concept class detection the student's data is divided into 10 chunks except the samples belonging to class 7 which are included after 8 chunk. Fig. 4. shows the prediction of student's data with concept class detection at chunk 8, as data of class 7 arrives after 8th chunk. It is shown in Fig. 4. with star mark.

For multiple concept class detection the dataset is divided into 10 chunks except the samples belonging to class 4 and class 5 which are included in 6th chunk and samples of class 1, 2, 3 are in first three chunks and samples of class 6 and 7 are in chunk 7 to 10. Fig. 4. shows the prediction of students data with multiple concept class detection with big circle at chunk 3, as data of class 4, 5 arrived in 6thchunk.

5. Conclusion

This paper intends to fill the breach between experimental classification of students for their career choice and the existing machine learning techniques of incremental learning concept. In a situation, where the data is being endlessly generated, storage of data is not possible for batch learning concept. Thus, incremental learning algorithm proposed in this paper, is found to be a useful technique for offering best career choice for the student. Use of incremental learning algorithm for the data of different types including time series, web log, spatial and multimedia is an important area of future work as this data is like a stream data, where the use of batch classifiers is impracticable. For combining voting, the strategy used in this paper is weighted majority vote, apart from this it may be significant to go for different mixture of rules to find the promptness between the rules having mixture, the individual classifier and the dataset used.

References

1. Hall JW. Adaptive Selection of U.S. Stocks with Neural Nets. *Trading on the Edge: Neural*, Genetic and Fuzzy System for Chaotic Financial Markets; 1994.
2. Cao LJ and Francis EH Tay. Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting. *IEEE Transactions on Neural Networks*; 2003.p. 1045-9227.
3. Christos Kyron, and Theocharis. Scope: Towards a Systolic Array of support Object Detection. *IEEE Embedded Systems Letters*; 2009. p. 1943-0663.
4. Arif S, Al-Hammadi and Milne RH. A neuro-fuzzy classification approach to the assessment of student performance. *IEEE conference Budapest Hungary*; 2004. p. 25-29.
5. Firat Hardalac. Classification of educational backgrounds of students using musical intelligence and perception with the help of genetic neural networks. *Expert systems with applications*; 2009. p. 6708-6713.
6. Millan JR. Rapid, safe, and incremental learning of navigation strategies. *IEEE Trans. Syst., Man, Cybern., Part B: Cybern*;1996. p. 408-420.
7. Chen GY, Tsai WH. A incremental-learning-by-navigation approach to vision-based autonomous land vehicle guidance in indoor environments using vertical line information and multiweighted generalized Hough transform technique. *IEEE Trans. Syst. Man, Cyern., Part B:Cybern*;1998. p.740-748,
8. Segal RB, Kephart JO. Incremental learning in swiftfile. In : Proc. 17th Int. Conf. Mach. Learn; 2000. p. 863-870.
9. Roozben Razavi-Far, Pieri Baraldi, and Enrico Zi. Dynamic weighting ensembles for incremental leaning and diagnosing new concept class faults in nuclear power system. *IEEE Trans. On Nuclear science*; 2012.
10. Christos Kyrkou, and Theocharis Theocharides. Scope: Towards a Systolic Array of SVM Object Detection. *IEEE Embedded Systems Letters*; 2009. p. 1943-0663.
11. Platt JC, Scholkopf B, Burges C. Fast Training of Support Vector Machines using Sequential Minimal Optimization. *Advances in kernel methods*, MIT Press; 1999. p.185-208,.
12. Vasilis A, Sotiris, Peter WT, and Michael GP. Anomaly Detection Through a Bayesian Support Vector Machine. *IEEE Transactions on Reliabi*. 2010. p. 0018-9529.
13. Peter GV Axelberg, Irene Yu-Hua Gu, HJ Bollen. Support Vector Machine for Classifications of Voltage Disturbances. *IEEE Transactions on Power* Delivery; 2007. 22-3.
14. Alistair Shilton, M. Palaniswami, Daniel Ralph, Ah Chung Tsoi. Incremental learning of Support Vector Machines. *IEEE Trans. On Neural Networks*; 2005. p. 114-131.
15. Zeki Erdem, Robi Polikar, Fikret Gurgen and Nejat Yumusak. Ensemble of SVMs for incremental learning. *Springer-verlag* Berlin Heidelberg; 2005. p. 246-256.
16. Nozaki K, Ishibuchi H, Tanaka H. Adaptive fuzzy rule based classifications systems. *IEEE Transactions Fuzzy System*; 1996. p. 238-250.
17. Yen GG, Meesad P. An effective neuro fuzzy paradigm for machinery condition health monitoring. *IEEE trans. Syst., Man, Cybern., Part B:Cybern*.; 2001. p. 523-536.
19. Arif S., Al-Hammadi, Milne RH. A neuro-fuzzy classification approach to the assessment of student performance. *IEEE conference Budapest Hungary*; 2004. p. 25-29.
20. Polikar R, Udpa L, Udpa SS, Honavar V. Learn++: An Incremental Learning Algorithm for Supervised Neural Networks. *IEEE Transactions on systems, Man and Cybernetics part C: Applications and review*; 2001.
21. Seiichi Ozawa, Shaoning Pang, Nikola Kasababov. Incremental learning of chunk data for online pattern classification systems. *IEEE Transactions on Neural Networks*; 2008. p. 1061-1074.
22. Muhlbaier M, Topalis A, Polikar R. Learn++.NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes. *IEEE Trans. Neural Netw*; 2009. p. 152–168.
23. Wolpert DH, Stacked generalization. *Neural Netw*; 1992. p. 241–259.
24. Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE. Adaptive mixtures of local experts. *Neural Comput*; 1991. p. 79–87.

25. Kotsiantis S, Kiriakos P, Xenos M. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based System*; 2010. p. 529-535.
26. Xu L, Kryzak A, Suen CY. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transaction on systems, Man and Cybernetics*; 1992. p. 418-435.
27. Jordan MI, Jacobs RA. Hierarchical mixtures of experts and the EM algorithm. *Neural Comput*; 1994. p. 181–214.
28. Kittler J, Hatef M, Duin RP, Matas J. On combining classifiers. *IEEE Trans. Pattern Anal. Machine Intell*; 1998. p. 226–239.
29. Rangarajan S, Jalote P, Tripathi S. Capacity of voting systems. *IEEE Trans. Software Eng*; 1993. p. 698–706.
30. Ji C, Ma S. Combination of weak classifiers. *IEEE Trans. Neural Netw*; 1997. p. 32–42.
31. He H, Yuan Cao. SSC: A Classifier combination method based on signal strength. *IEEE Trans. On Neural Networks and learning system;* 2012.
32. Ade R, Deshmukh PR. Classification of students using psychometric tests with the help of incremental naïve bayes algorithm. *IJCA*; 2014. p.27-31.
33. Ade R, Deshmukh PR. An incremental ensemble of classifiers as a technique for prediction of student's career choice. *IEEE Conference on ICNSC*; 2014. p. 427-430.
34. Gunduz G, Fokoue E. UCI Machine Learning Repository. Available http://archive.ics.uci.edu/ml/datasets/ Irvine, CA: University of California, School of Information and Computer Science; 2013.